

BAB I

PENDAHULUAN

1.1 Latar Belakang

World Health Organization (WHO), pada tanggal 30 Januari 2020 mengumumkan penyakit corona virus (COVID-19) sebagai penyakit yang menyebar infeksi. Virus ini mula-mula muncul di Cina pada tahun Desember 2019, dan sejak itu menyebar dengan cepat di seluruh dunia, berdampak negatif pada sejumlah negara di Eropa, Amerika, dan Asia [1].

Dalam upaya untuk menahan penyebaran COVID-19, di sebagian besar negara di seluruh dunia, lembaga pendidikan telah memutuskan untuk sementara menangguhkan pengajaran tatap muka dan beralih ke model pembelajaran jarak jauh. Hal ini menyebabkan banyak negara-negara, baik eropa dan asia melakukan lockdown guna membatasi kunjungan dan aktifitas. Kebijakan pemerintah Indonesia untuk meminimalis terjangkitnya wabah covid-19, mengeluarkan perintah bahwa kegiatan-kegiatan seperti bekerja, belajar dan beribadah dapat dilakukan dari rumah, ini dilakukan sebagai bentuk tanggung jawab pemerintah terhadap kesehatan dan keselamatan rakyat. Dampak yang di timbulkan dari pembelajaran during terhdap pelajar adalah dampak psikologis yaitu mental.

Dengan berlakunya pendidikan jarak jauh menimbulkan dampak negatif terutama dikalangan pelajar, terdapat beberapa perubahan yang dialami oleh siswa di saat menjalani pendidikan jarak jauh diantaranya melalui empat saluran utama : waktu yang dihabiskan untuk belajar lebih sedikit, gejala stress, perubahan cara siswa berinteraksi, dan kurangnya motivasi belajar [2].

Stres dapat digambarkan sebagai suatu keadaan dimana seseorang mengalami gangguan jiwa karena kesulitan/perubahan yang sedang atau dialaminya [3]. Menurut [4] Stres adalah pikiran atau perasaan yang muncul sebagai respon terhadap kesulitan atau ancaman yang disebut sebagai stressor.

Sangat dibutuhkan solusi dalam mengatasi berbagai kendala untuk pembelajaran jarak jauh agar siswa tetap dapat belajar dengan baik. Analisa yang baik terhadap permasalahan maupun kendala yang dihadapi akan memudahkan dalam mencari

solusi yang tepat. Machine learning memegang peranan penting dalam menganalisa dan mengolah data, salah satu metode yang dilakukan adalah dengan pengklasifikasian, metode ini mampu mengelompokkan atau mengategorikan beberapa data yang diatur secara berurutan. Di dalam kehidupan sehari-hari kasus klasifikasi sering di hadapi, baik itu pengelompokan data pada bidang pendidikan, sosial, instansi terkait, dan pada sektor-sektor tertentu lainnya. Munculnya kendala pengelompokan terhadap beberapa instansi-intansi terkait di karenakan belum tepatnya dalam penggunaan metode klasifikasi yang diinginkan, sehingga dapat mempengaruhi dalam menghitung akurasi dari algoritma yang diterapkan. Klasifikasi pada data dalam jumlah banyak dan dengan fitur atau atribut yang beragam seringkali membuat hasil akurasi menjadi rendah. Untuk itu diperlukan metode yang dapat mengatasi data dengan jenis beragam tersebut.

Penggunaan metode klasifikasi data seperti : Decision Tree, K-Nearest Neighbors, Random Forest, *Naïve Bayes*, Logistic Regression, dan algoritma lainnya merupakan kategori algoritma yang sering digunakan untuk melakukan pengelompokan atau klasifikasi data. Pada penulisan ini, penulis akan membahas penerapan metode yang digunakan yaitu klasifikasi Logistic Regression dan *Naïve Bayes*.

Logistic Regression adalah model klasifikasi statistik probabilistik yang memprediksi probabilitas terjadinya dari sebuah acara. Logistic Regression memodelkan hubungan antara variabel dependen kategoris X dan a hasil kategoris dikotomis atau fitur Y [5]. Logistic Regression memodelkan hubungan antara kategorikal variabel dependen dan out- kategoris dikotomis datang atau fitur. Ini digunakan sebagai model biner (ganda) untuk memprediksi respons biner (banyak), hasil dari variabel dependen egoris, berdasarkan satu atau lebih variabel independen [6]

Selain Logistic Regression, klasifikasi dapat dilakukan juga dengan menggunakan metode *Naïve Bayes*. *Naïve Bayes* adalah dasar lain yang diawasi model pembelajaran mesin berdasarkan teorema Bayes, dengan asumsi independensi observasi. Keuntungan utama *Naïve Bayes* bahwa model pembelajarannya mudah dan tidak memerlukan perhitungan yang rumit dari iterative parameter [7].

Penggunaan metode Logistic Regression dan Naïve Bayes pada proses klasifikasi data memerlukan metode tambahan untuk mendapatkan performa dan keakuratan terbaik, diperlukan pemilihan fitur agar relevansi setiap fitur dengan keluaran dapat diukur, oleh karena itu penggunaan Feature Selection menjadikan kinerja metode akan lebih baik, salah satunya adalah Forward Selection

Pada penelitian sebelumnya yang berjudul “*Comparative Analysis of Stress among Undergraduate Students Using Random Forest and Logistic Regression Techniques*” [8]. Hasil analisis dan implementasi data berjalan dengan baik, memberikan akurasi 95,23% dan skor uji f1 sebesar 0,9411 menggunakan regresi logistic dan akurasi = 90,47% dan skor f1 = 0,8888 menggunakan Random Forest. Penelitian lain dengan judul “*Mental Stress Detection in University Students using Machine Learning Algorithms*” [9]. Dengan hasil : Random Forest : sensitivity =100 , specivity = 66.66 dan akurasi = 83.33 , Naïve Bayes : sensitivity = 66.66, specivity = 75 dan akurasi = 71 Support Vector Machine sensitivity = 75, specivity= 100 , akurasi = 85.71 4, K-Nearest Neighbors sensitivity = 70, specivity = 44, akurasi = 55.55. Penelitian berikutnya dengan judul “*Komparasi Algoritma K-Nearest Neighbor dan Algoritma Naïve Bayes terhadap Evaluasi Pembelajaran Daring*” [10]. Hasil ketepatan Algoritma Naïve Bayes dengan angka 91, 45%. Hasil ketepatan Algoritma K- Nearest Neighbor dengan angka 97, 72%.

Berdasarkan uraian diatas, penulisan ini nantinya akan menghasilkan, nilai akurasi perbandingan dari klasifikasi algoritma *Naïve Bayes* , Logistic Regression , dan penggunaan optimalisasi Feature Forward Selection pada algoritma Logistic Regression (LR) dan *Naïve Bayes (NB)* berdasarkan akurasi, *error*, *sensitivity*, dan AUC (under area curve) sebagai nilai akhirnya, dengan kategori-kategori yang tersedia didalam data dampak stress pelajar di masa pandemi covid-19 yang terdapat pada dataset public.

Dengan menggunakan metode ini, di harapkan hasil penelitian ini akan mampu memberikan nilai akurasi terbaiknya, baik sebelum menggunakan optimalisasi feature forward selection maupun sesudah menggunakan optimalisasi feature forward selection.

1.2. Ruang Lingkup

Ruang Lingkup pada penelitian ini adalah sebagai berikut:

1. Penerapan metode yang digunakan pada penelitian ini adalah *Naïve Bayes* dan Logistic Regression dalam menentukan hasil nilai akurasi dari klasifikasi Dampak Stres pelajar di Masa Pandemi Covid-19.
2. Optimalisasi yang digunakan dalam meningkatkan performa metode Logistic Regression dan *Naïve Bayes* adalah Feature Forward Selection
3. Data yang diambil adalah data public tentang dampak stress pelajar di masa Pandemi Covid-19 yang telah tersediapada alamat <https://www.kaggle.com/datasets/kunal28chaturvedi/covid19-and-its-impact-on-students>

1.3 Rumusan Masalah

Dengan dasar pemikiran yang tertera pada latar belakang penulisan, maka dapat diambil rumusan masalah sebagai berikut :

1. Bagaimanakah perbandingan nilai akurasi sebelum menggunakan Feature Forward selection dari algoritma *Naïve Bayes* dan Logistic Regression terhadap klasifikasi dampak stress pelajar di masa Pandemi Covid-19?
2. Bagaimanakah penerapan optimalisasi Feature Forward Selection dalam menyeleksi variable pada *Naïve Bayes* dan Logistic Regression ?
3. Bagaimanakah pengaruh optimalisasi setelah menggunakan Feature Forward selection pada algoritma *Naïve Bayes* dan Logistic Regression untuk klasifikasi Dampak Stres pelajar di Masa Pandemi Covid-19?

1.4 Tujuan Penelitian

Rumusan masalah pada penelitian ini adalah bertujuan sebagai berikut :

1. Menentukan nilai akurasi sebelum menggunakan Feature Forward Selection dari Algoritma *Naïve Bayes* dan Logistic Regression pada klasifikasi dampak stress pelajar di masa Pandemi Covid-19
2. Mengkaji dan menyeleksi variable pada penerapan optimalisasi Feature Forward Selection guna meningkatkan performa pada Logistic Regression dan *Naïve Bayes*

3. Membandingkan nilai akurasi setelah menggunakan Feature Forward Selection dari Algoritma *Naïve Bayes* dan Logistic Regression serta menganalisis pengaruh optimasi Feature Forward Selection pada algoritma Logistic Regression dan *Naïve Bayes* berdasarkan *error*, *sensitivity*, *akurasi* dan AUC (under area curve) untuk klasifikasi Dampak Stres pelajar di Masa Pandemi Covid-19.

1.5 Manfaat Penelitian

Dari penelitian ini, yang dapat diambil manfaat penelitiannya adalah sebagai berikut :

1. Dapat menambah pemahaman dan mengembangkan wawasan dari metode data mining seperti *Naïve Bayes*. dan Logistic Regression
2. Menambah keilmuan dari metode optimasi Feature Forward Selection dalam meningkatkan performa pada metode data mining.
3. Menambah pengetahuan dan dapat memberikan motivasi sebagai pembanding dan saran bagi peningkatan penelitian untuk metode selanjutnya yang akan dilakukan.

1.6 Sistematika Penulisan

Untuk mempermudah pembahasan dan pemahaman maka penulis membuat sistematika penulisan sebagai berikut :

BAB 1 PENDAHULUAN

Dalam bab ini berisikan mengenai latar belakang perumusan masalah, ruang lingkup penelitian, tujuan penelitian dan manfaat penelitian

BAB 2 TINJAUAN PUSTAKA

Dalam bab ini menjelaskan mengenai uraian secara terstruktur mengenai teori-teori pendukung yang berkaitan dengan kasus yang digunakan untuk merancang bangun system yang berhubungan dengan penelitian ini.

BAB 3 METODE PENELITIAN

Pada bab ini akan dibahas tahapan objek penelitian, jenis penelitian, sumber data, waktu penelitian, penggunaan variabel dan perancangan diagram alir atau work flow.

BAB 4 HASIL DAN PEMBAHASAN

Bab ini menerangkan mengenai mekanisme proses pengujian terhadap data yang digunakan pada penelitian ini. Analisa yang dilakukan meliputi nilai akurasi dari hasil klasifikasi algoritma yang diterapkan berikut dengan nilai perbandingan dari masing-masing algoritma dengan beberapa referensi yang akan dipaparkan.

BAB 5 SIMPULAN DAN SARAN

Pada bab penutup ini akan menerangkan isi dari kesimpulan yang di ambil dari hasil penelitian dan pengujian yang telah dilakukan. Masukan berupa saran dan kritik yang sifat nya membangun diharapkan untuk pengembangan lebih lanjut dituliskan juga pada bab ini.

DAFTAR PUSTAKA

Referensi atau daftar pustaka berisi jurnal ilmiah, hasil penelitian orang lain, buku – buku, dan bahan – bahan yang dapat dijadikan sebagai refrensi dalam pembahasan karya tulis.

LAMPIRAN

Bagian ini berisi data yang dapat mendukung atau memperjelas pembahasan atau uraian yang dikemukakan dalam bab – bab sebelum